

(51) Int. Cl. 7
 H04R 3/00
 G10L 15/00
 21/02
 15/20

識別記号
 320

F I
 H04R 3/00
 G10L 3/00
 3/02
 301
 301

7-マコ-7 (参考)

5D015
 J 5D020
 F
 E

審査請求 未請求 請求項の数2 O L (全6頁)

(21) 出願番号 特願平11-121518

(22) 出願日 平成11年4月28日 (1999.4.28)

(71) 出願人 000101732

アルバイン株式会社

東京都品川区西五反田1丁目1番8号

(72) 発明者 齋藤 望

東京都品川区西五反田1丁目1番8号 アルバイン株式会社内

(72) 発明者 中田 孝一

東京都品川区西五反田1丁目1番8号 アルバイン株式会社内

(74) 代理人 100084711

弁理士 齋藤 千幹

最終頁に続く

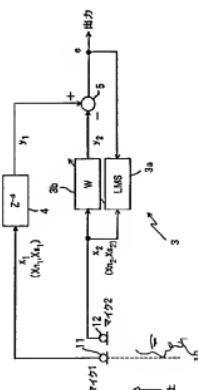
(54) 【発明の名称】マイクロホンシステム

(57) 【要約】

【課題】 マイクロホンを2つ使用するマイクロホンシステムにおいて、音声信号のSN比を改善する。

【解決手段】 2つマイクロホン11, 12の出力信号を用いて適応信号処理を行いSN比を改善した話者音声信号を出力するマイクロホンシステムにおいて、各マイクロホン11, 12を接近して配置すると共に、一方のマイクロホンから出力する信号のSN比を高くし、他方のマイクロホンから出力する信号のSN比を低くする。例えば、一方のマイクロホン11を話者10の顎の真上に配置し、他方のマイクロホン12を該一方のマイクロホンの位置より1~5cm程度後頭部側に離して配置する。

本発明のマイクロホンシステムの構成



【特許請求の範囲】

【請求項1】 2つマイクロホンの出力信号を用いて適応信号処理を行いSN比を改善した話者音声信号を出力するマイクロホンシステムにおいて、

各マイクロホンを接近して配置すると共に、一方のマイクロホンから出力する信号のSN比を高くし、他方のマイクロホンから出力する信号のSN比を低くすることを特徴とするマイクロホンシステム。

【請求項2】 前記一方のマイクロホンを話者の顔の真上に配置し、他方のマイクロホンを該一方のマイクロホンの位置より1~5cm程度後頭部側に離して配置することを特徴とする請求項1記載のマイクロホンシステム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明はマイクロホンシステムに係り、特に、2つマイクロホンの出力信号を用いて適応信号処理を行いSN比を改善した話者音声信号を出力するマイクロホンシステムに関する。

【0002】

【從来の技術】 現在の音声認識システムは、15dB以上のSNR(S:音声/ノ:ノイズ)が確保されている場合、約95%の認識率を実現できるくらいの技術レベルにまで達している。しかし、周囲に存在するノイズによりSN比が低下すると、それに伴って認識率が急激に低下する性質も有している。図8はSN比と認識性能との関係をいくつかの種類のマイクロホン(無指向性、單一指向性、狭指向性、AMNR(AdaptiveMicrophone-array for Noise Reduction))について評価したもので、SN比と認識率がおおむねS字特性100を示す帶の中に含まれている。この図8から明らかなように、認識率はSN比の低下により急激に低下し、SN比が0dBの環境下において約50%まで低下してしまう。

【0003】 そのため、自動車が発生するノイズ(エンジン音、ロードノイズ、バーンノイズ、風切り音など)が存在する自動車車室内において、上記のような認識性能の劣化は避けられず、音声認識システムを車載化する上で大きな問題の一つとなっている。前記したような事情から、周囲に存在するノイズの影響を少なくし、高いSN比で音声を受音するための方式が種々提案されており、複数のマイクロホンとデジタル信号処理を用いた高SN比受音システムはその一例である。かかる高SN比受音システムの中で最も簡単な構成のものは図9に示すように2つのマイクロホンを使用するシステムであるが、他にも、Griffith-Jim型アレイやAMNRといった、より高度なシステムが提案されている。

【0004】 図9において、1、2は第1、第2のマイクロホン、3は適応信号処理部であり、誤差信号eが入力されると共にマイクロホン2の出力信号x₂が参照信号として入力され、誤差信号eのパワーが最小となるようにLMS(Least Mean Square)アルゴリズムに基づいて適

応信号処理を行う。適応信号処理部3において、3aはLMS演算部、3bは例えばFIR型デジタルフィルタ構成の適応フィルタである。LMS演算部3aは適応信号処理により誤差信号eのパワーが最小となるように適応フィルタ3bの係数を決定する。

【0005】 4はマイクロホン1から出力されるノイズ信号を目標信号として入力される目標応答設定部であり、音響系の逆特性を精度よく近似するためのものである。適応フィルタ3bのタップ長の半分の信号遅延時間tをdとするとき、目標応答設定部4は該時間dの遅延特性を有し、オーディオ周波数帯域でフラットな特性(ゲイン1の特性)を有する。すなわち、目標応答設定部4は、図10(a)に示すようにゲイン1のフラットな周波数特性を備え、図10(b)に示すように遅延時間dを有するインバ尔斯応答特性を有している。この目標応答設定部4は、FIR型デジタルフィルタの遅延時間dに応する係数を1にし、他の係数を0にすることにより実現できる。5は減算部であり、目標応答設定部4から出力する目標応答より適応フィルタ3bの出力信号を減算して誤差信号eを出力する。

【0006】 非音声認識時、マイクロホン1、2にはノイズのみが入力し、適応信号処理部3は適応信号処理により誤差信号e、すなわち、ノイズ出力のパワーが最小となるようにフィルタ係数Wを決定する。一方、音声認識時には、適応信号処理部3はフィルタ係数の更新をせず、前記非音声認識時に決定したフィルタ係数Wを適応フィルタ3bに設定して音声信号を出力する。図9に示すシステムに本来求められている理想的な性能は、音声認識時に出力信号として音声信号x_s(z)のみ(ノイズ出力0)を出力することである。すなわち、ノイズ出力E_n(z)に関する、

$$E_n(z) = X_{n1}(z)z^{-4} - X_{n2}(z)W(z) \quad (1)$$

$\{E_n(z)\}^2$ の平均が最小値をとるよう調整可能なパラメータ(適応フィルタ3bの係数)Wを決定することである。ただし、X_{n1}(z)、X_{n2}(z)はマイクロホン1、2の出力信号に含まれるノイズ信号である。

【0007】

【発明が解決しようとする課題】 自動車には騒音源が多数存在するため、マイクロホン1、2が拾う自動車車室内ノイズのコーヒレンズは、マイクロホン1、2を遠ざけるにしたがって低下する傾向を有している。このため、2つのマイクロホン1、2を遠ざける程、(1)式の条件が満たされにくくなってしまう問題が生じ。マイクロホン1、2はできるだけ近い位置に配置する必要がある。ところが、2つのマイクロホン1、2をできるだけ近い位置に配置すると、2つのマイクロホン1、2には同様の音声とノイズがそれぞれ入射する可能性が高くなり、(1)式を満たすように適応フィルタ係数Wを決定してノイズを消去すると、音声までも消去してしまう。一方、音声が歪まないように適応フィルタ係数Wを決定す

ると、ノイズがほとんど消えず、SN比もほとんど改善されなくなってしまうという問題が発生する。

【0008】かかる問題は、図9に示したシステム特有のものではなく、Griffith-Jim型アレイやAMNRといった、より高度な高SN比受音システムを採用した場合でも、ほぼ同様に発生する。以上から本発明の目的は、マイクロホンを2つ使用する図9のマイクロホンシステムにおいて、音声信号のSN比を改善できるようにすることである。

【0009】

【課題を解決するための手段】上記課題は本発明によれば、2つマイクロホンの出力信号を用いて適応信号処理を行いSN比を改善した話者音声信号を出力するマイクロホンシステムにおいて、各マイクロホンを接近して配置すると共に、一方のマイクロホンから出力する信号のSN比を高くし、他方のマイクロホンから出力する信号のSN比を低くすることにより達成される。このようにすれば、2つのマイクロホンの出力信号に含まれるノイズ $X_{n1}(z)$, $X_{n2}(z)$ を略等しくでき、一方、2つのマイクロホンの出力信号に含まれる音声信号 $X_{s1}(z)$, $X_{s2}(z)$ を異ならせることができる。従って、ノイズ信号入力時の $E_n(z)$ の2乗平均値が最小となるように適応フィルタ係数 W を決定しても、(2)式の音声出力 $E_s(z)$ は零とならず、音声信号のSN比を改善することができる。

【0010】マイクロホンの具体的な配置例としては、一方のマイクロホンを話者の顔の真上に配置し、他方のマイクロホンを該一方のマイクロホンの位置より1~5cm程度後頭部側に離して配置する。このようにすれば、人間の音声放射特性により、マイクロホンが比較的近距離に配置されているにも拘らず、1つのマイクロホンはできるだけ高いSN比で音声を拾い、もう一方のマイクロホンでは、できるだけ低いSN比で音声を拾うようになります。

【0011】

【発明の実施の形態】(a) マイクロホンシステムの構成

図1は本発明のマイクロホンシステムの構成図であり、図9のシステムと同一部分には同一符号を付している。図中、1, 0は話者であり、例えば自動車の運転手、1, 1, 2は第1、第2のマイクロホンである。第1のマイクロホン1, 1は話者の顔の真上天井に配置し、第2のマイクロホン1, 2は第1のマイクロホン位置より1~5cm程度後頭部側の天井に配置する。3は適応信号処理部で、誤差信号 e が入力されると共にマイクロホン2の出力信号 x_s が参照信号として入力され、誤差信号 e のパワーが最小となるようにLMSアルゴリズムに基づいて適応信号処理を行う。適応信号処理部3において、3aはLMS演算部、3bはFIR型デジタルフィルタ構成の適応フィルタである。LMS演算部3aは適応信号処理により誤差信号 e のパワーが最小となるように適応フィルタ3b

の係数を決定する。適応信号処理部3は、非音声認識時においてのみ適応信号処理により適応フィルタ3bのフィルタ係数 W を決定し、音声認識時にはフィルタ係数の更新をせず、前記非音声認識時に決定したフィルタ係数 W を適応フィルタ3bに設定する。4はマイクロホン1, 1から出力する信号を目標信号として入力される目標応答設定部で、時間 d の遅延特性を有し、かつ、オーディオ周波数帯域でフラットな特性(ゲイン1の特性)を有している。5は減算部で、目標応答設定部4から出力する目標応答より適応フィルタ3bの出力信号を減算して誤差信号 e を出力する。

【0012】(b) 人間の音声放射特性

図2は人間の音声放射特性であり、(a)は話者1の0の口元を含む水平面において口元から所定距離の位置における音声レベルを周波数毎に示す放射特性図、(b)は話者1の0の口元を含む直面において口元から所定距離の位置における音声レベルを周波数毎に示す放射特性図である。図中、Aは125Hz~250Hz、Bは500Hz~700Hz、Cは1400Hz~2000Hz、Dは4000Hz~5600Hzの特性である。この放射特性図より明らかのように、人間が発生する音声は、話者正面方向に最も強く放射され、上方や下方、及び左右方向に放射される音声のパワーは、話者正面方向に比べて小さい。それゆえ、図1のように第1のマイクロホン1, 1を話者の顔の真上天井に配置し、第2のマイクロホン1, 2を第1のマイクロホン位置より1~5cm程度後頭部側の天井に配置すれば、①2つのマイクロホン1, 1, 1, 2で受音するノイズのパワーをほぼ同一にできる一方、②2つのマイクロホン1, 1, 1, 2で受音する音声パワーを異ならせることができる。すなわち、2つのマイクロホン1, 1, 1, 2の出力信号に含まれるノイズ $X_{n1}(z)$, $X_{n2}(z)$ を略等しくできると共に、2つのマイクロホン1, 1, 1, 2の出力信号に含まれる音声信号 $X_{s1}(z)$, $X_{s2}(z)$ を異なせることができ、「 $[X_{n1}(z)/X_{n2}(z)] \neq [X_{s1}(z)/X_{s2}(z)]$ 」とすることが可能である。

【0013】(c) 動作

マイクロホン1, 1, 1, 2にノイズのみが入力する非音声認識時において、適応信号処理部3は適応信号処理により次式

$$E_n(z) = X_{n1}(z)^{-4} - X_{n2}(z)W(z) \quad (1)$$

において、「 $E_n(z)$ 」の平均値が最小となるように適応フィルタ3bのフィルタ係数 W を決定する。一方、音声認識時、適応信号処理部3はフィルタ係数の更新をせず、前記非音声認識時に決定したフィルタ係数 W を適応フィルタ3bに設定して音声信号を出力する。この場合、マイクロホン1, 1, 1, 2の出力に含まれる音声信号 $X_{s1}(z)$, $X_{s2}(z)$ は異なり、「 $[X_{n1}(z)/X_{n2}(z)] \neq [X_{s1}(z)/X_{s2}(z)]$ 」となるため、次式

$$E_s(z) = X_{s1}(z)^{-4} - X_{s2}(z)W(z) \quad (2)$$

により求まる音声出力 $E_s(z)$ は最小値には(ノイズと異なり、あまり小さくはない)ならない。以上より、(1)式のノイ

ズ出力 $E_n(z)$ のパワーが0となるように適応フィルタ係数 W を決定しても、(2)式の音声出力 $E_s(z)$ はノイズと同様に小さくはならず、音声信号のSN比を改善することができる。

【0014】以上要約すれば、図1に示すように、マイクロホン1, 1, 12を比較的近距離に配置することで2つのマイクロホンが输出するノイズ間のコヒーレンスの低下を少なくし、更に、図2に示すような人間の音声放射特性を考慮することで、マイクロホン1, 1, 12が比較的近距離に配置されているにも拘らず、1つのマイクロホン1はできるだけ高いSN比で音声を拾い、もう一方のマイクロホン12では、できるだけ低いSN比で音声を拾う。この結果、ノイズ出力が零となるように適応フィルタ係数 W を決定しても、音声出力はノイズと同様に小さくはならず、音声信号のSN比を改善することができる。

【0015】(d) マイクロホン位置とSN比改善量の検討

図2の放射特性より人間が発声し空間に放射した音声は、後頭部側で特に大きく減衰し、正面に放射される音声に比べてレベルが小さくなることがわかる。それゆえ、本発明のマイクロホンシステムは、図1に示したように人間の頭部真上付近から後頭部側にかけて設置するのが基本であり、このように第1、第2のマイクロホン1, 1, 12を設置することによりSN比をより大きく改善することができる。図3はペアマイクロホン位置の説明図、図4は図3の各ペアマイクロホン位置におけるSN比改善量を示す図である。図3に示すように、3cm間隔のペアマイクロホン1, 1, 12を複数の位置①、②、③に設置し、それぞれのSN比の改善量(どの程度SN比が向上するか)を、1500ccの乗用車(セダン)で調べてみると、図4に示す結果が得られた。この図4より、ペアマイクロホン1, 1, 12を①の位置に設置したとき、すなわち、「1つのマイクロホンを話者10の頭のほぼ真上に設置し、もう1つのマイクロホンを少し離して後頭部側に配置したとき」にSN比の改善量が最も高くなることがわかる。

【0016】図5はペアマイクロホン間隔の説明図、図6は図5の各ペアマイクロホン間隔におけるSN比改善量を示す図である。図5に示すように、第1のマイクロホン1を話者10の頭のほぼ真上に固定し、第2のマイクロホン12を後頭部側にそれぞれ3cm, 6cm, 9cm, 12cm離して配置し、最適なマイクロホンの間隔について調べてみると、図6に示す結果が得られた。この図6より、2つのマイクロホン1, 1, 12の間隔が狭いほど、SN比の改善量は高いと考えられる。しかし、図1に示したシステムでは、間隔を0cmとするとノイズを完全に消去できるが、音声もまた完全に消去してしまう。このため、音声受音システムとして機能しないことになる。また、小型マイクロホンといえどもそれ自体の大きさ

があるため、マイクロホンどうしを完全にくっつけても、マイクロホンの中心間隔は1cm程度より小さくはならない。それゆえ、マイクロホンの間隔は、車種の違いとマイクロホンの大きさにより若干の幅があるものの、せいぜい1~5cm程度にするのがよい。

【0017】図7は発声者別のSN比改善量説明図である。この図7より明らかのように、本発明のマイクロホンシステムでは、人の違いによる性能(SN比改善量)のばらつきは1dB程度であり、話者の違いによる影響は少ない。以上2つのマイクロホンを話者の頭上に配置した場合について説明したが、2つのマイクロホンを比較的近距離に配置し、1つのマイクロホンにより、できるだけ高いSN比で音声を拾い、もう一方のマイクロホンにより、できるだけ低いSN比で音声を拾うようにできれば、配置位置は頭上に限らない。以上、本発明を実施例により説明したが、本発明は請求の範囲に記載した本発明の主旨に従い種々の変形が可能であり、本発明はこれらを排除するものではない。

【0018】

【発明の効果】以上本発明によれば、各マイクロホンを接近して配置すると共に、一方のマイクロホンから出力する信号のSN比を高くし、他方のマイクロホンから出力する信号のSN比を低くするようにしたから、ノイズ出力が最もとなるように適応フィルタ係数を決定しても、音声出力は零とならず、音声信号のSN比を改善することができる。すなわち、少ないマイクロホン数であるにも拘らず、高いSN比で音声を受音出力することができる。又、本発明によれば、一方のマイクロホンを話者の頭の真上天井に配置し、他方のマイクロホンを該一方のマイクロホンの位置より1~5cm程度後頭部側に離して配置することにより、マイクロホンが比較的近距離に配置されているにも拘らず、1つのマイクロホンはできるだけ高いSN比で音声を拾い、もう一方のマイクロホンでは、できるだけ低いSN比で音声を拾うようにできる。

【図面の簡単な説明】

【図1】本発明のマイクロホンシステムの構成図である。

【図2】人間の音声放射特性図である。

【図3】ペアマイクロホンの位置説明図である。

【図4】ペアマイクロホン位置とSN比改善量の関係図である。

【図5】ペアマイクロホンの間隔説明図である。

【図6】ペアマイクロホンの間隔とSN比改善量の関係図である。

【図7】発声者別SN比改善量説明図である。

【図8】SN比と認識率の関係図である。

【図9】従来のマイクロホンを2つ使用した場合の高SN比受音システムである。

【図10】目標応答設定部の特性図である。

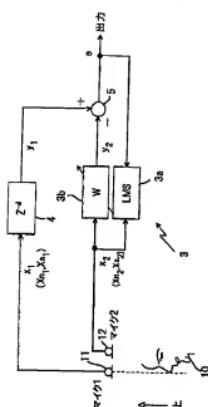
【符号の説明】

1 1, 1 2 . . 第1、第2のマイクロホン
 3 . . 適応信号処理部
 3 a . . LMS演算部

3 b . . 適応フィルタ
 4 . . 目標応答設定部
 5 . . 減算部

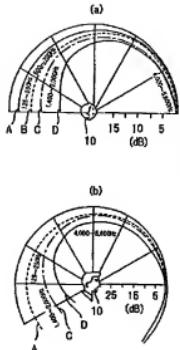
【図 1】

本発明のマイクロホンシステムの構成



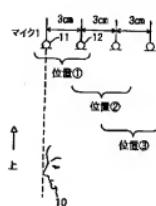
【図 2】

人間の音声放射特性



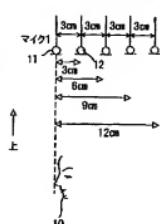
【図 3】

ペアマイクロホンの位置説明図

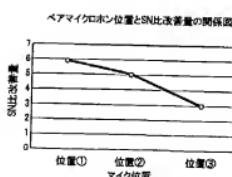


【図 5】

ペアマイクロホンの間隔説明図

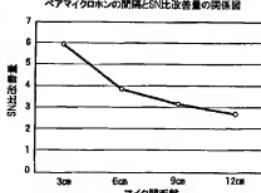


【図 4】



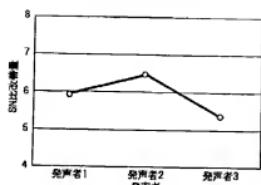
【図 6】

ペアマイクロホンの間隔とSN比改善量の関係図



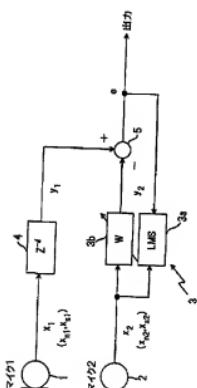
【図 7】

免声者別SN比改善量



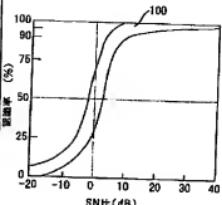
【図 9】

従来のマイクロホンを2つ使用した場合の高SN比受音システム



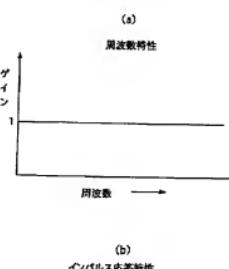
【図 8】

SN比と認識率の関係



【図 10】

目標応答設定部の特性



フロントページの続き

(72)発明者 木内 真吾

東京都品川区西五反田1丁目1番8号 ア
ルバイン株式会社内

Fターム(参考) 5D015 CC14 EE05 KK02

5D020 BB07